Forecasting Research

Forecasting Research Division
Technical Report No. 226

Performance of the Nimrod System

by

B W Golding

November 1997

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

DUPLICATE

NATIONAL METEOROLOGICAL LIBRARY

Forecasting Research
Technical Report No. 226

Perfomance of the Nimrod System

by

B W Golding

November 1997

Forecasting Research
Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

© Crown Copyright 1997

Permission to quote from this paper should be obtained from the above Meteorological Office division.

Performance of the Nimrod System

by

B.W.Golding

November 1997

1. Objectives

The Project set out to determine and implement the optimum method of generating analysis and forecast guidance in the time range 0-6 hours for the UK and immediately surrounding areas, and the optimum method of producing cloud/rainfall fields for use in initialising the mesoscale model.

Development of improvements to the mesoscale NWP model was specifically excluded. It was accepted that the project was dependent on improvements to the quality of model products.

There was a presumption that the method selected would be fully automatic in order to meet cost and timeliness constraints.

2. Product Specification

Introduction

All products are generated and stored centrally in gridded form, as values representative of a grid area or volume. The derivation of site specific forecasts was not part of the project. All products can be made available in polar stereographic and national grid coordinate systems.

Reliability

Reliability is measured in two ways: internally by logging completion of the rainfall forecast script, and externally by receipt of rainfall forecasts at EA(Thames). Relative to the first, the latter measure rejects successful forecasts when there are telecommunications problems, but accepts degraded forecasts

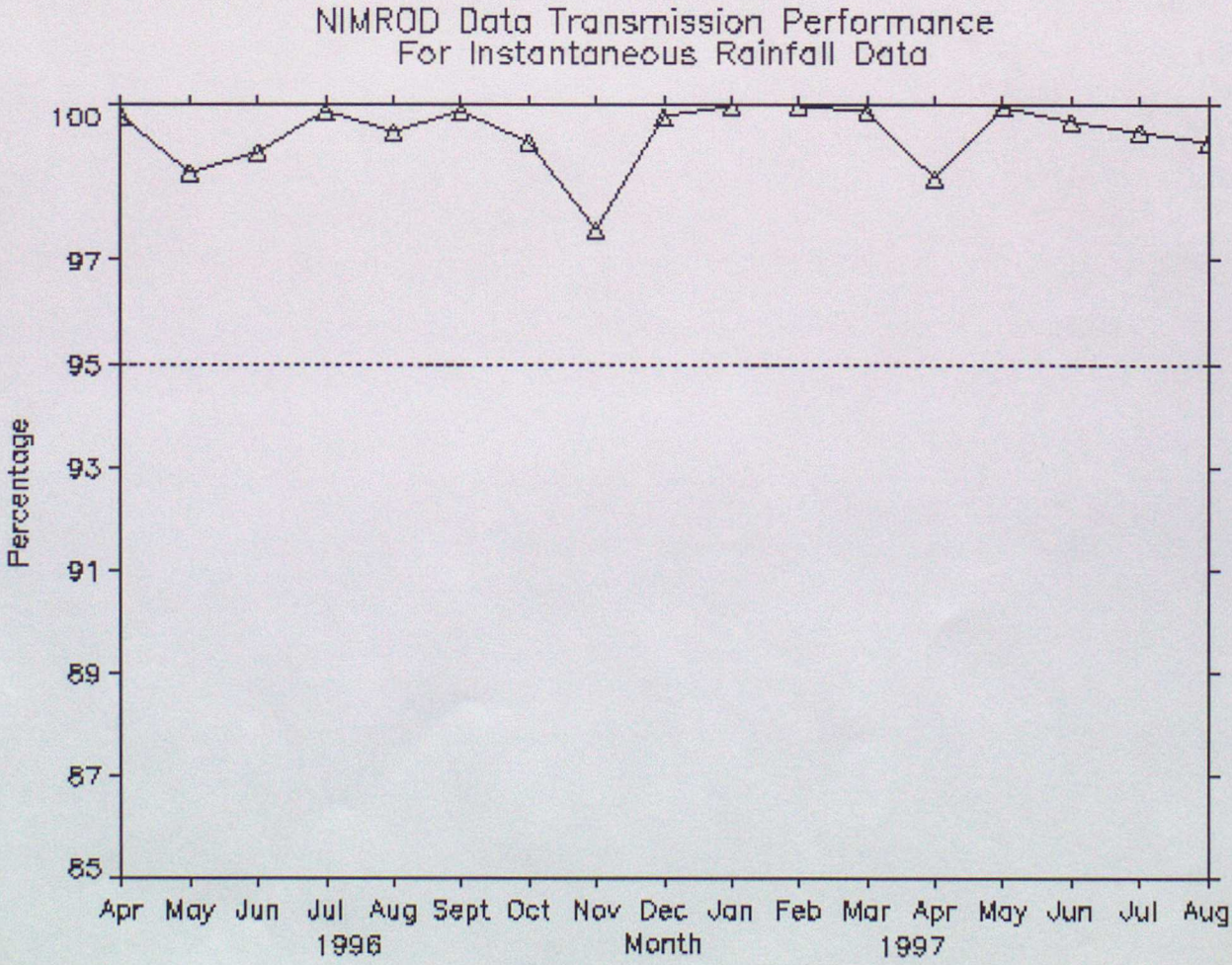

Fig.1 Nimrod reliability measured at EA(Thames)

produced without NWP input following switchover to the backup processor. Fig. 1 shows results from the EA(Thames) analysis indicating that the mandatory availability threshold of 95% has been achieved continuously since January 1996, and the target (98%) in all months except January and November 1996.

There is no record of mean outage. However, system management is set up to enable recovery, either on the main processor or on the backup, to be achieved with the loss of only one forecast cycle (i.e. 1 hour at most), so it is believed that the requirement for a mean outage of less than 2 hours is satisfied. Longer outages are only likely in the event of an external cause such as failure of the CDN.

Although current reliability is excellent and timeliness requirements are largely being met, there is concern that modest increases in load, such as occur during archiving and when access to the monitor is made, have an excessive impact on performance. Such increases caused the reduction in reliability noted in November 1996 due to the impact of the visibility analysis & forecast. Evidence suggests that this is due to Nimrod operating close to maximum load much of the time. It is likely to be exacerbated by the addition of extra tasks required to meet new contract requirements and to deliver the remaining core products to NMC

Variables

Nimrod currently generates five precipitation variables, seven cloud variables, four visibility variables and one wind variable. The precipitation related variables are: location/intensity, accumulation, precipitation type, probability of snow, lightning rate & CAPE. The cloud related variables are cloud amount below 1000', cloud amount below 5000', total cloud amount, cloud base, cloud base greater than 2 oktas, cloud base greater than 4 oktas, and cloud top. The four visibility related variables are: mean visibility, "worst" visibility, screen temperature and screen dew point. The wind variable is 10m gust speed. These meet most, but not all, of the products in the current version of the User Requirement for Nimrod. The missing variables are 10m and 50m wind vector, which can be obtained from the mesoscale model when required, and probability of thick fog, which is currently under development.

It should be noted that the original User Requirement contained many additional variables that have been removed mainly because there was little likelihood of the users being able to receive and display the products on the time scale of the project. It is anticipated that some may be requested at a later date. The original customer for each product is given in brackets as defence (D), commercial (C) or NMC (N). They are:

Snow depth -available from the mesoscale model; can easily be provided in Nimrod format (D,C,N)
Convection areas - could provide product similar to the area of non-zero lightning probability (D,C,N)
Liquid water content - could provide meso. model modified for consistency with Nimrod cloud f/c (C,N).
Sunshine - could provide product based on total cloud cover (C).
2m wind - could be diagnosed from mesoscale model output (C).
300m wind - could be diagnosed from mesoscale model output (D,C,N).
ground temperature - available from mesoscale model (C)
300m temperature - could be diagnosed from mesoscale model (C,N).
850mb temperature - available from mesoscale model (C,N).
comfort index - could be diagnosed from existing Nimrod products (C).
fog point - could be diagnosed from existing Nimrod products (N).
wind chill index - could be diagnosed from existing Nimrod products (C).
freezing level - could be diagnosed from mesoscale model (D,C,N).
wet bulb freezing level - available (D,C,N).
msl pressure / pressure tendency - available from mesoscale model (C,N).

Delivery

Only precipitation rate and type are currently available to outstations, the first on ODS and the second on OPUS. The remainder are unlikely to be available at outstations until NIMBUS is implemented. NMC currently receives precipitation rate & type, cloud cover & base>4oktas, and mean visibility. Further products will be delivered as HORACE acquires the capability to receive them. The COSMOS image database currently receives precipitation rate & type. Further products will be delivered as customer requirements develop, subject to the capability of the database and of communications between Nimrod and COSMOS.

Quality

The quality assessment for individual products is given in the following sections. The User Requirement recognised that areally defined variables could not validly be compared with point observations. The option of correcting such comparisons using a measure of the representativeness of observations was rejected in favour of stating the acceptance criteria in terms of verification against analyses. Where there are qualitative differences in the results obtained against point observations, these are referred to in the relevant section. Many of the error measures displayed in this paper are available on the intranet via the Nimrod monitor.

Two levels of accuracy were defined in the User Requirement. For all of the main variables, mandatory acceptance criteria were specified. These were considered to provide a guide to the level of performance below which operational production would not be worthwhile. Where possible target criteria were also set which would significantly enhance product usefulness. Due to lack of experience in the quality of some variables, target criteria were only set for precipitation and wind (also temperature and pressure, which were subsequently removed from the UR). While providing useful justification for developing improvements beyond the mandatory quality, the target criteria for precipitation are not realistic. This is illustrated by considering the error measures obtained by making a small displacement to a typical rainfall analysis (see Golding, Met. Apps.,1997, Nimrod: a system for generating automated very short range forecasts). The target criteria for precipitation location at T+1,3 corresponded to displacements of about 5,12 km while those for precipitation rate at T+1,3,6 corresponded to displacements of 4,6,7 km. Clearly, given an initial uncertainty in position arising from the 5km radar resolution, it is unrealistic to aim for such a high level of accuracy. Both criteria are referred to in the relevant sections and are marked on the verification graphs.

3. Precipitation Rate & Accumulation

Forecast Cycle & Interval

Precipitation rate and accumulation forecasts are output every 15mins to 6 hours ahead from each forecast, as required. However, forecasts are updated every 30mins rather than the 15min cycle requested. Most customers do not have the means to receive data from a 15min cycle, so this shortcoming is not considered a high priority at present.

Domain & Resolution

Precipitation rate & accumulation are provided on a 5km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid and reprojected to a Polar Stereographic projection. All specifications are satisfied.

Timeliness

The requirement was for precipitation analyses and forecasts to be available by 15mins after data time. The following table gives availability times for September 1997 (the most recent month available) relative to nominal time: Whereas the analysis is available well within target, recent additions to the forecast processing have delayed its completion. A priority for the new hardware will be to meet this timeliness target.

	Average	95% available
Radar composite	7.61min	
Precipitation rate analysis	8.40min	9.00min
Precipitation forecasts	16.46min	21.00min

Accuracy of precipitation location

Forecast locations of precipitation rate exceeding 0.125mm.hr^{-1} and 10mm.hr^{-1} are verified, both at full resolution and after smoothing with a 25km half-width Gaussian. The requirement was expressed in terms of the lower threshold after smoothing. Results for the latest 12 month period are shown in figure 2. The mandatory hit rate is exceeded at all times, and substantially so at T+1 and T+6. During the past year, both mandatory criteria have been met in 6 months at T+1 (best performance 85% HR / 24% FAR); in 8 months at T+3 (best performance 71% HR / 40% FAR); and in 6 months at T+6 (best performance 67% HR / 55% FAR). The missing months at T+1 were dry, with the verifying actuals dominated by isolated pixels, both real and spurious.

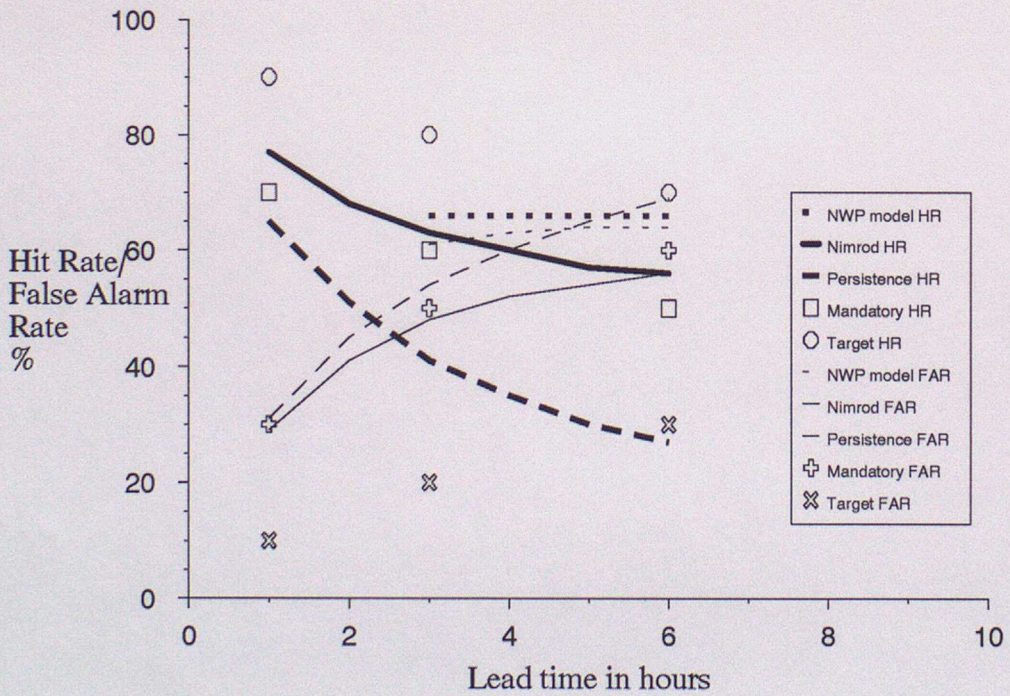

Fig. 2 Accuracy of smoothed areas of rainfall rate > 0.125 mm/hr
September 1996 - September 1997.

Figures have also been computed for the accuracy of intensities exceeding 10mm/hr. However, since the mean scale of such rain areas is lower than the smoothing scale, it is only possible to verify at full pixel resolution. Bearing in mind the uncertainty due to timing & aliasing errors, and contamination of the actual by spurious echoes, it is not surprising that the results are poor. Over the past year, the hit rate at this threshold has been 8% for pure advection and 6% for the Nimrod forecast. The useful

message from this assessment is that pure advection should be given greater weight for intense precipitation in the early part of the forecast. An upgrade to achieve this was implemented in 3/97.

Accuracy of shower distribution

It was requested that the temporal and spatial distribution of showers should be correctly represented on at least 50% of occasions when classified as isolated (>100km), occasional (50-100km) and frequent (<50km). No objective measure of this criterion has been attempted. Subjectively, the shower structure is well maintained to T+1 but has been largely lost by T+6. Upgrades to improve this aspect of performance were implemented in 3/97. An inevitable side effect is a slight worsening of other verification measures.

Accuracy of precipitation intensity

The key measure (revised following preliminary results) is the accuracy of 15km averaged rates against analyses where either forecast or analysis exceed 0.125 mm hr^{-1}. A minimum value of half the threshold is imposed. Results for the latest 12 month period are shown in figure 3. The mandatory requirement is met at T+6. Note that the model, if available as quickly as Nimrod rather than at T+3, would outperform pure advection by T+2 and persistence by T+1. During the past year, the mandatory criterion was met at T+1 in 4 months (best performance 3.56), at T+3 in 2 months (best performance 4.92) and at T+6 in 8 months (best performance 5.07).

Accuracy of precipitation accumulation

The key measure (revised following preliminary results) is the accuracy of 15km averaged one hour accumulations against analyses where either forecast or analysis exceed 0.2mm. A minimum value of

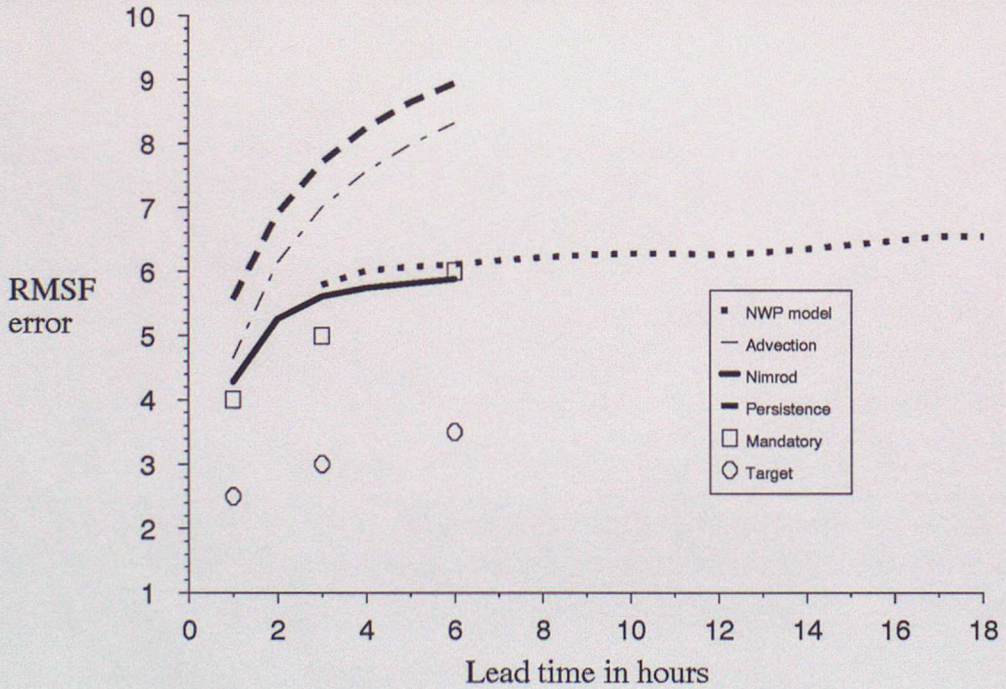

Fig. 3 Fractional errors in 15km average rainfall rate forecasts
September 1996 - September 1997

half the threshold is imposed. Results for the latest 12 month period are shown in figure 4. The mandatory requirement is met at T+1 and nearly so at T+6. It should be noted that use of averaged 15minute rates to estimate the hourly actual accumulations introduces a significant uncertainty in this comparison which will be much reduced when 5 minute radar data become available. During the past year, the mandatory criterion was met at T+1 in 7 months (best performance 2.19), at T+3 in no months (best performance 3.84) and at T+6 in 3 months (best performance 4.15).

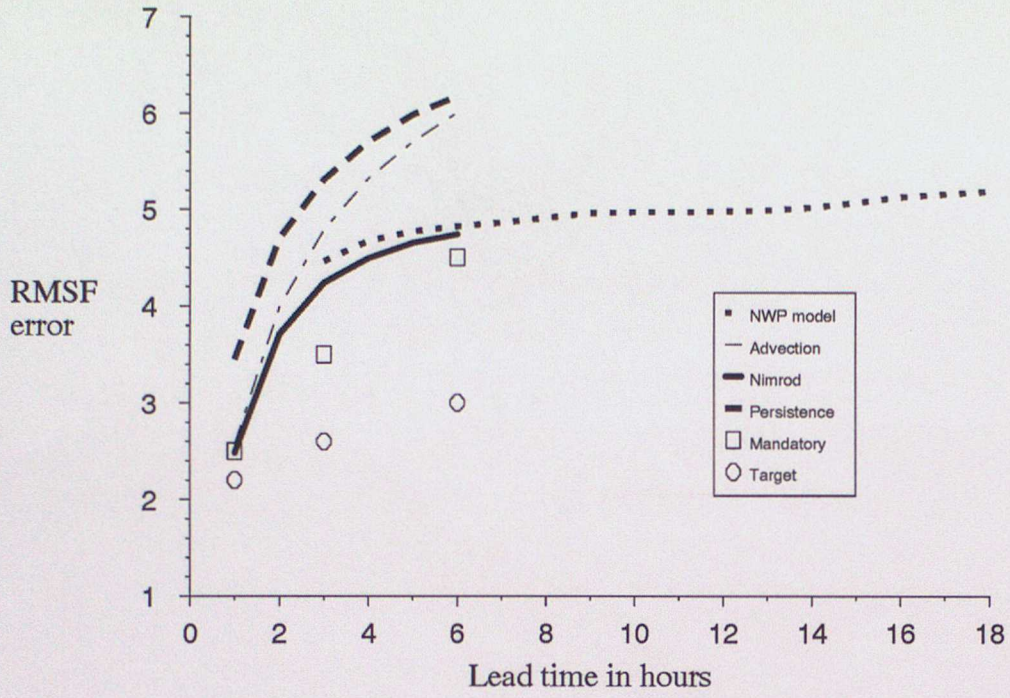

Fig.4 Fractional error in 15km average rainfall accumulation forecasts
September 1996 - September 1997

Accuracy of analyses vs raingauges

Verification against raingauge reports has only been carried out for estimates of actual hourly rainfall accumulation. Table 1 shows various measures of accuracy computed for two of these estimates. The first is obtained from a weighted average of the rain rate at each pixel from the radar composite. If a radar is missing for part of the hour, the remaining images are used to define the mean rain rate. The second estimate is obtained from a weighted average of rainfall analyses. These are identical to the composite over areas within the computed "area of radar coverage", but obtained from a combination of surface reports, satellite derived rain and model forecast outside. Comparisons are made where either Nimrod or the gauge exceeds the threshold. A minimum value of 0.1mm is imposed for both thresholds to avoid undefined ratios in the fractional statistics. The RMS and RMSF errors have decreased significantly over the past year, probably due to improved radar performance.

	Radar composite		Analysis	
	>0.2mm	>1.0mm	>0.2mm	>1.0mm
Mean	-0.05	-0.13	-0.03	-0.10
RMS	1.28	2.13	1.27	2.11
MeanF	1.00	0.94	1.07	1.00
RMSF	3.52	4.26	3.44	4.10
% within x2	39%	47%	39%	47%

Table 1 Accuracy of hourly rainfall accumulation actuals compared with raingauges
September 1996 - September 1997

The two estimates are similar, the analysis having slightly smaller errors. The bias is negligible, both additively and multiplicatively. There is an uncertainty in the gauge reports of small amounts since the timing of a tip of the gauge is not directly related to the time of occurrence of that amount of rainfall. This is probably the reason for a greater percentage of comparisons, at the higher threshold, being within a factor of two. The greater RMSF error at this threshold seems contradictory, but indicates that those cases with greater fractional error are dominated by very large errors - some of these may be small positional errors but most are probably due to spurious echoes.

Overall, these errors are unacceptably large. However, it is not clear how much of the difference is due to spatial representativeness errors, where for most applications the radar can be considered the more useful measure, how much to temporal representativeness, which is purely a function of the radar image frequency, how much to spurious echoes, and how much to genuine errors in the retrieval of rain rate from the radar. With the routine availability and processing of 2km, 5minute images from some radars, it will be possible to substantially reduce the representativeness errors, allowing a better understanding of where improvements need to be made.

Subjective assessment of precipitation rate

The trial of Nimrod vs FRONTIERS in August & September 1995 concluded that Nimrod was superior to FRONTIERS in all respects except spurious echo removal. This has remained a problem, with numerous complaints, particularly related to presentation quality, but also including some events where genuine showers were removed. The User Requirement has been taken to be for a product which has a negligible probability of removing real precipitation. Absence of cloud is probably the only totally reliable indicator, since dry surface reports, alone, do not indicate absence of precipitation in nearby areas. However, improved performance has been achieved, especially in daylight with the use of visible satellite imagery. A revised scheme, implemented in November 1996, appears to have removed the worst problems. It remains the case that spurious echoes are retained if there are surface precipitation reports in the area and it is cloudy. A probability of precipitation field is diagnosed to enable precipitation echoes to be distinguished from spurious ones. This field could be used, with different thresholds, to meet different requirements for the reliability of echoes in the precipitation actual.

4. Precipitation Type

Types

The following types are computed, as defined in the user requirement: rain, wet snow, dry snow, hail, freezing rain, sleet, drizzle, freezing drizzle. A requirement for a rime product has recently been added for commercial aviation and defence users. A riming rate will be computed, using the visibility, wind speed and temperature products, for a 1cm diameter cable. The raw product will be in $kg.m^{-1}.hr^{-1}$ but it is anticipated that it will be presented to users as slight, moderate or severe.

Forecast Cycle & Interval

Precipitation type forecasts are output every hour to 6 hours ahead and are updated every hour.

Domain & Resolution

Precipitation type diagnoses and forecasts are provided on a 15km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid and reprojected to a Polar Stereographic projection.

Timeliness

Precipitation type products were available by T+17 on average and T+21 on 95% of occasions during September 1997.

Accuracy

The type is determined almost entirely by the mesoscale model fields, only hail diagnosis depending significantly on the Nimrod rain rate prediction. For this reason, verification has concentrated on the

Nimrod T+0 diagnosis, which maximises the correlation between observed and forecast precipitation areas. Comparison is with surface synoptic observations, and is only carried out when precipitation of some sort is both observed and diagnosed. The range of mesoscale model forecasts that is effectively being verified is T+3 - T+9. Table 2 is a contingency table with forecast types in columns and observed types in rows. Unfortunately, the verification statistics have been contaminated by a processing error which mis-assigns observed "smoke" to freezing rain and "haze" to snow. This invalidates the statistics for freezing rain and causes a significant lowering of the scores for snow, especially in summer. The following characteristics of the quality of precipitation type diagnoses are drawn from the uncontaminated statistics and from subjective analysis of results:

(i) Snow is well diagnosed. During significant snow events, the diagnosis performs very well, achieving hit rates of over 90%. However, it is less successful at capturing isolated snow reports, especially showers. The figures for the full year are distorted by the verification error.

(ii) Hail diagnoses are subjectively realistic, but the scale of actual hail occurrences is small in time and space, making verification difficult. In particular, the errors are often related to the time between synoptic observations and the nominal time.

(iii) Freezing rain is not well forecast except on occasions when the classical temperature profile for its occurrence is well marked in neighbouring radiosonde ascents assimilated by the mesoscale model.

(iv) Drizzle is very poorly diagnosed.

%	f/c drizzle	f/c rain	f/c mixed snow/rain	f/c freezing rain/dz	f/c snow	f/c hail
obs drizzle	1.3%	7.5%	0.2%	0.0%	0.2%	0.0%
obs rain	2.8%	76.6%	1.7%	0.0%	0.9%	0.3%
obs mixed snow/rain	0.1%	1.1%	0.3%	0.0%	0.4%	0.0%
obs freezing rain/dz	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%
obs snow	0.7%	1.9%	0.3%	0.0%	3.0%	0.0%
obs hail	0.0%	0.3%	0.1%	0.0%	0.1%	0.0%

Table 2 Accuracy of diagnosed precipitation type compared with surface reports
January - September 1997

5. Snow probability

This product is not a project requirement, but is generated as part of the precipitation type algorithm, and for some purposes may be superior.

Forecast Cycle & Interval

Snow probability forecasts are output every hour to 6 hours ahead and are updated every hour.

Domain & Resolution

Snow probability diagnoses and forecasts are provided on a 5km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid.

Timeliness

Snow probability products are generated at the same time as precipitation type and were available by T+17 on average and T+21 on 95% of occasions during September 1997.

Accuracy

No verification of this product has been undertaken, but the skill of snow diagnosis in the precipitation type product indicates that it performs well. Subjectively there is a much clearer correlation with altitude than is apparent in the precipitation type product offering, for instance, useful guidance on the level above which snow can be expected.

6. Lightning Rate

The lightning products currently disseminated to customers, provide a measure of lightning risk. The Nimrod product is a strike rate. However, it correlates well with the probability of observing a flash, and it is expected that the product will be used in the form of a lightning risk.

Forecast Cycle & Interval

Lightning rate forecasts are output every hour to 6 hours ahead from each forecast, and are updated every hour.

Domain & Resolution

Lightning rate is provided on a 15km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid.

Timeliness

Lightning rate products are generated at the same time as precipitation type and were available by T+17 on average and T+21 on 95% of occasions during September 1997.

Accuracy

Lightning rate is verified against ATD fixes. However, the ATD system does not record all lightning strikes. It is understood that, on average, about 10% of strikes over the UK are recorded. It can not be assumed that every active storm is identified. The system is also subject to positional errors, which increase from about 10km over the UK to much larger values over Europe and the Atlantic.

The output variable from Nimrod is the strike rate per minute per 15km pixel. In order to estimate a verifying strike rate, it is necessary to accumulate ATD reports over a period. The longer the period, the more reliable the estimated strike rate, but the greater the contamination by advection or development. A 30 minute period is currently used. The verification has changed twice since May 1996 when its current form was instituted. In September, the forecast field was smoothed to attempt to allow for advection over the half hour period. This improved the hit rate at the expense of false alarm rate. From January 1997, the verification was restricted to the UK land area, to avoid the observation errors associated with more distant ATD fixes, and the diagnosis errors associated with unreliable precipitation rates outside radar cover. It also corresponds more directly with the user requirement.

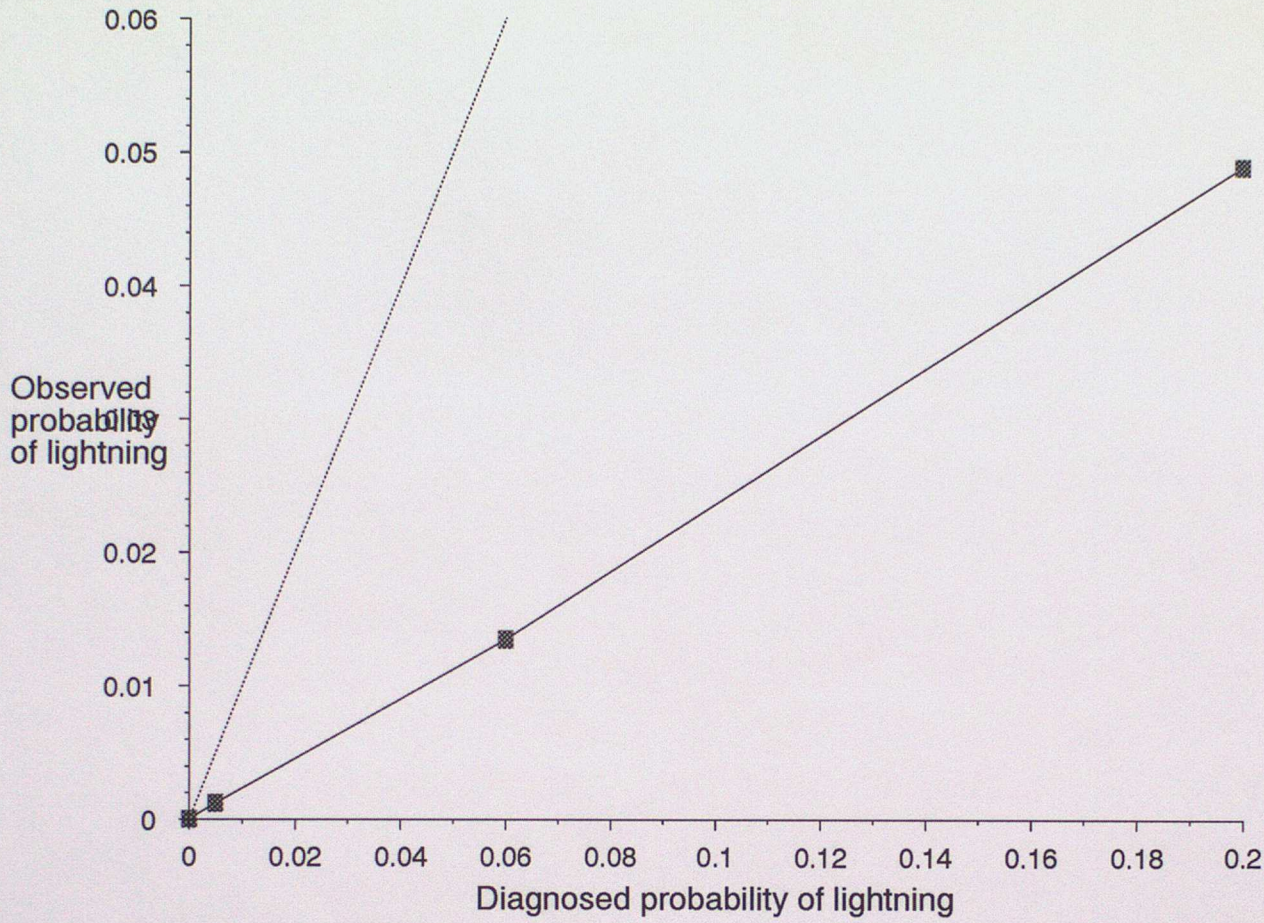

Fig.5 Diagnosed vs observed lightning probability, January - September 1997.
(dotted line denotes perfect agreement)

Fig.5 shows that the forecast probabilities are strongly related to the observed ones, but with substantially higher values. These are due to a combination of under-reading by the ATD observation system, and over-diagnosis by Nimrod. Taking a 1% probability threshold gives a hit rate of 93.7%. Allowing a 50km margin of error, increases this to 99.3%, but with a 96.7% false alarm rate. These results suggest that a useful product could be contructed with lightning risk categories as follows:

Low	Within 50km of 1% diagnosed probability or more
Medium	1% diagnosed probability or more
High	6% diagnosed probability or more

7. Cloud

Forecast Cycle & Interval

Cloud forecasts are output every hour to 6 hours ahead from forecasts updated every hour. Output every 30mins was requested and could be supplied if required. Otherwise all requirements are met.

Domain & Resolution

Cloud products are provided on a 15km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid and some products are reprojected to a Polar Stereographic projection. All specifications are satisfied.

Levels

The cloud analysis and forecasts are performed at the following heights above mean sea level in metres: 30, 60, 90, 120, 150, 180, 210, 250, 300, 360, 430, 520, 650, 800, 1000, 1200, 1500, 2200, 3000, 4000, 5000, 7000, 9000, 11000, 15000, 21000. This is significantly better resolution than was

requested at low levels, but slightly worse at high levels, where 5000' resolution was requested up to 45000'.

Timeliness

The requirement was for cloud analyses and forecasts to be available by 25mins after data time. They were available by T+23 on average and T+28 on 95% of occasions in May 1997.

Accuracy of location of cloud below 1000'

This is the most demanding of the specified requirements for accuracy of cloud. While forecasts of extensive stratus sheets, in this height range, are of high accuracy (though not as good as persistence forecasts), the majority of observations of more than half cloud cover below 1000' are isolated or short-lived, being related to local uplift or convection, and are not representative of a 15km pixel. In this situation verification against smoothed analyses may not be realistic due to the problem of smoothing these small areas, but verification against observations will also have a major representativeness contribution.

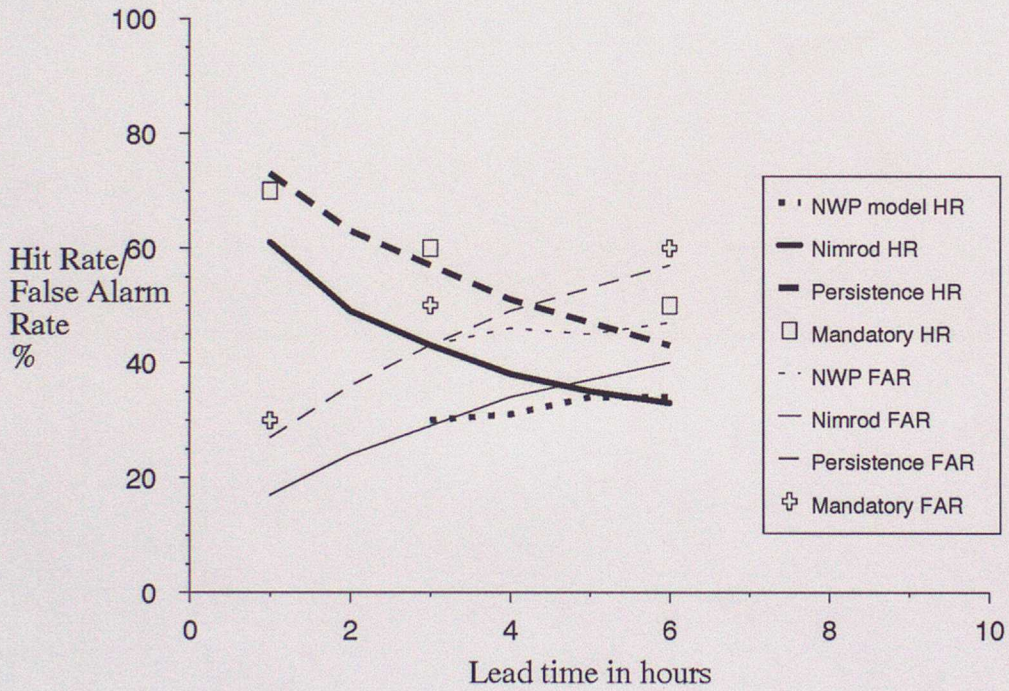

Fig. 5 Accuracy of smoothed areas of cloud below 1000ft > 4 oktas vs analyses September 1996 - September 1997.

Figure 5 shows the verification against smoothed analyses. The mesoscale model underpredicts systematically, and shows no skill, with the false alarm rate substantially higher than the hit rate at all verification times. The model underprediction pulls down the Nimrod hit rate and false alarm rate from those obtained by persistence, but hardly alters the crossover point. However, taken at face value, persistence is the better forecast and nearly meets the mandatory criteria. A correction to remove the model bias has been developed and will be trialled at the earliest opportunity. With such a correction it should be possible for Nimrod to meet the mandatory criteria.

Over the past year, the false alarm rate criteria have been met by Nimrod at all forecast times in all months. However, the hit rate criterion has not been met at T+3 (best performance 56%) & T+6 (best performance 46%) and has only been met twice at T+1 (Dec 96 & Jan 97, when there were periods with extensive stratus sheets. Best performance 73%).

Accuracy of location of cloud below 5000'

This measure is much less demanding than the last, covering the full range of boundary layer cloud. More than half cover in this range occurs frequently over large areas. Because of this the criterion was changed to focus on the gaps, but these are also reasonably frequent and large scale. The weakness of this measure is the definition of truth, since there are no reports of this quantity. Rather, it must be computed from cloud layer reports, making assumptions about overlap. The perennial problem of defining cloud cover is also present, especially with deep convective cloud which may be reported as

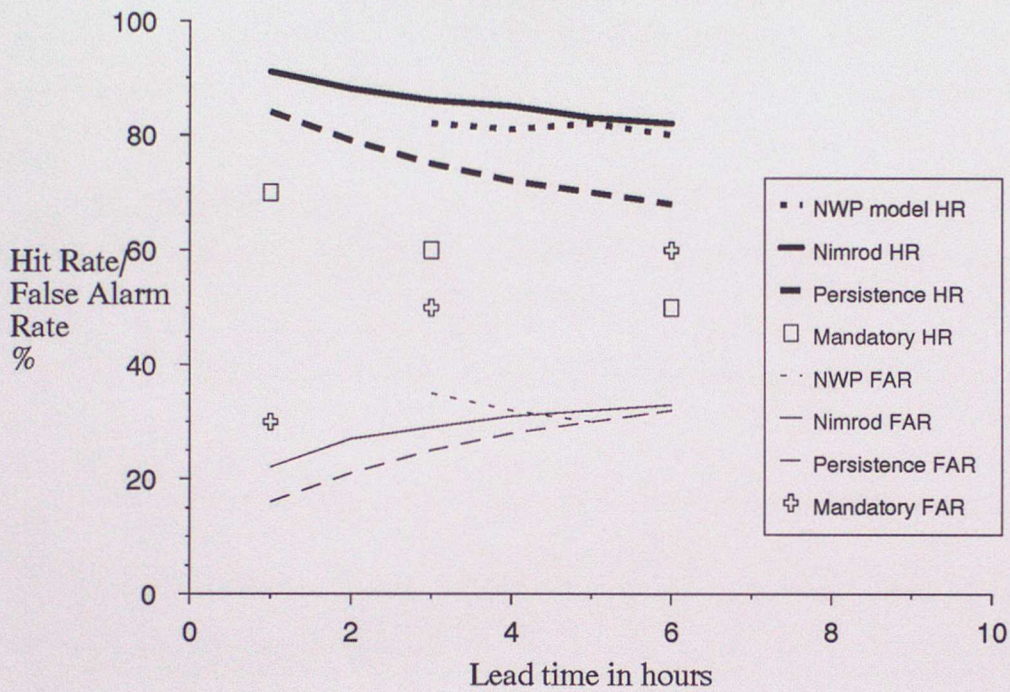

Fig. 6 Accuracy of smoothed areas of cloud below 5000ft < 4 oktas vs analyses September 1996 - September 1997.

anything from one to eight oktas at a particular point while filling the same fraction of a grid square. When incorporated in the analysis, it is inevitable that a significant amount of background field information will be preserved, which may affect the results on some occasions.

The results in figure 6 show that all forecast methods comfortably satisfy the mandatory criteria, with Nimrod substantially the best. This is also the case for the original criteria relating to more than half cover.

Accuracy of location of cloud at all levels

Apart from the problem of the definition of cloud cover in the presence of deep convective clouds, this is the most reliable of the cloud criteria, being well observed by satellite and from the ground. As with the last section, the criterion was changed to verify areas of less than half cover since these are much less common, and therefore more difficult to forecast, than areas of greater than half cover.

The results in figure 7 show that all forecast methods meet the mandatory criteria, though not by as much as for cloud below 5000'. Nimrod is worse than persistence at T+1 and better at T+6, but the differences are small. The model is slightly worse than these.

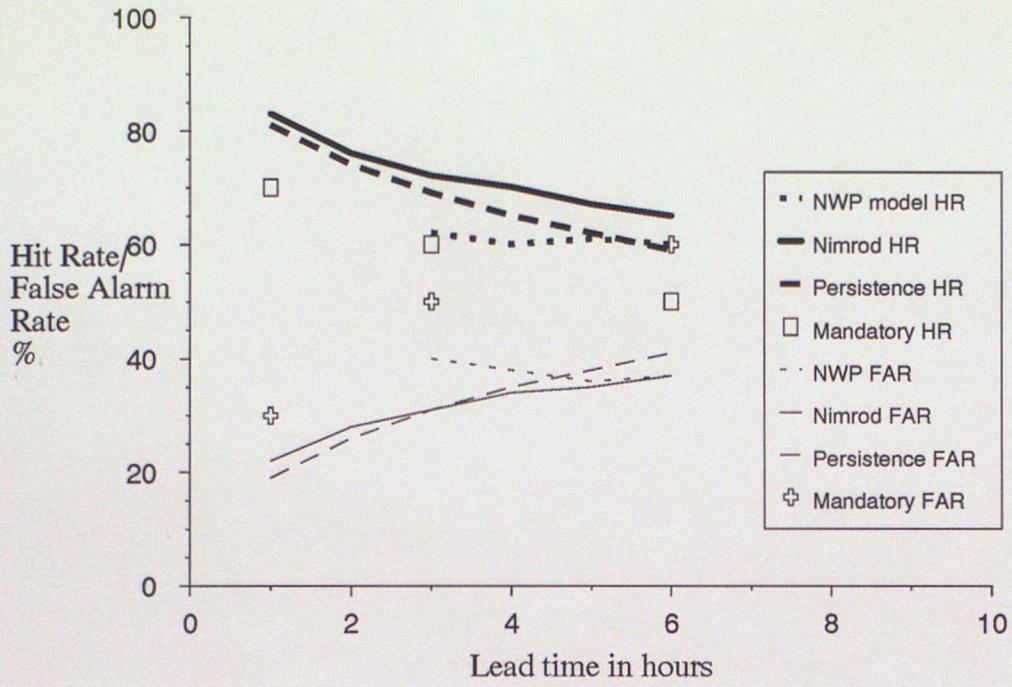

Fig. 7 Accuracy of smoothed areas of total cloud < 4 oktas vs analyses
September 1996 - September 1997.

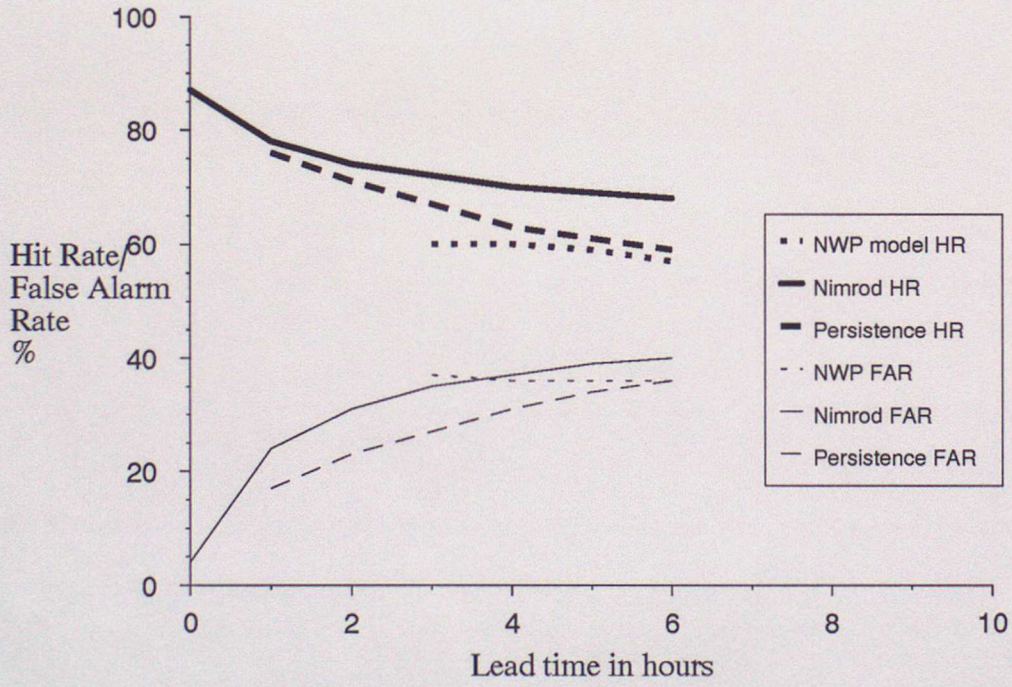

Fig. 8 Accuracy of total cloud < 4 oktas vs observations
September 1996 - September 1997.

A characteristic of some of the cloud scores, of which this is one, is that the performance measured against analyses is worse than that measured against point observations, despite the representativeness problems associated with the latter. This may reflect a weakness in the analysis, though this would tend to affect the forecasts as well, but is more likely a reflection of the impact of smoothing. Figure 8 shows the total cloud verification computed against point observations for the same period as Figure 7. It also shows the analysis fit to observations at T+0. Using this measure, Nimrod is clearly an improvement on both persistence and the model.

Accuracy of cloud base height

No criterion for accuracy of cloud base was specified, but it is an important aspect of the forecasts, perhaps the most important for RAF forecasting. As a guideline, we note the ICAO requirement for TAFs, that 75% of forecasts should be within the greater of 30m or 30% of actual.

Figure 9 shows results for the past year, measured against analyses, with no smoothing. A comparison is only made where either forecast or analysis is below 1500m. Values exceeding this are capped at 3000m. The mesoscale model is markedly worse than persistence or Nimrod. This is largely due to the bias noted earlier. As noted before, a bias correction technique has been devised, which should result in further improvement to the Nimrod product.

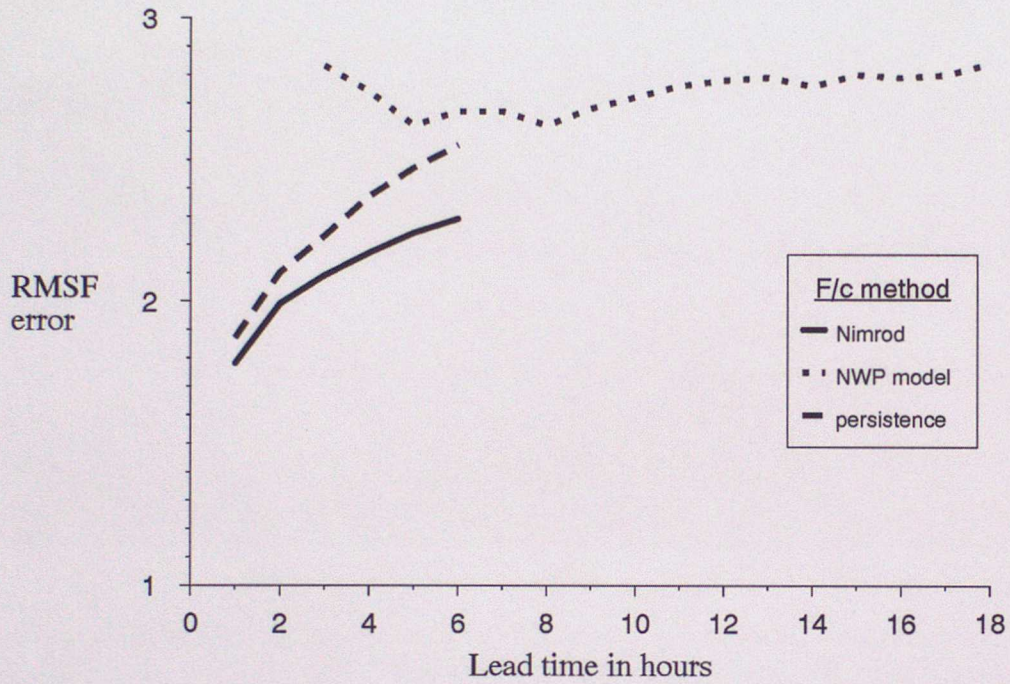

Figure 9 Fractional errors in 3 okta cloud base forecasts vs analysis
September 1996 - September 1997

In comparison with the precipitation and visibility scores, these are small errors, which also grow slowly, even for persistence. This is unlikely to be a result of using analyses as truth since site specific verification of forecasts gives RMSF errors only slightly (about 10%) larger

In terms of the ICAO requirement, the percentage within the required fractional error decreases from 51% (66%) at T+1 to 31% (41%) at T+6 for comparisons against observations (analyses).

Accuracy of total cloud cover

No requirement was specified in the project for this measure. However, it adds to the information provided by the measure of location accuracy for total cloud, and is a measure routinely computed for the mesoscale model. It is also well observed, although the difference between point observation and areal coverage is a problem for isolated patches of low cloud, such as thunderstorms.

Errors are computed at full resolution (15km) and so are affected by aliasing errors. However, the analysis uses a smoothing filter, so it is unlikely that errors at the individual pixel scale are contaminating the results to any significant degree.

Nimrod performs well relative to persistence and the model, particularly at T+3-4. However, the general level of error is higher than one would wish. Subjectively, a feature of the forecasts is the trend towards

middle values, resulting in clear areas being replaced with 1-2 oktas and overcast areas by 6-7 oktas. On most occasions this does not, however, compromise the value of the forecast when used to generate a four category aviation forecast.

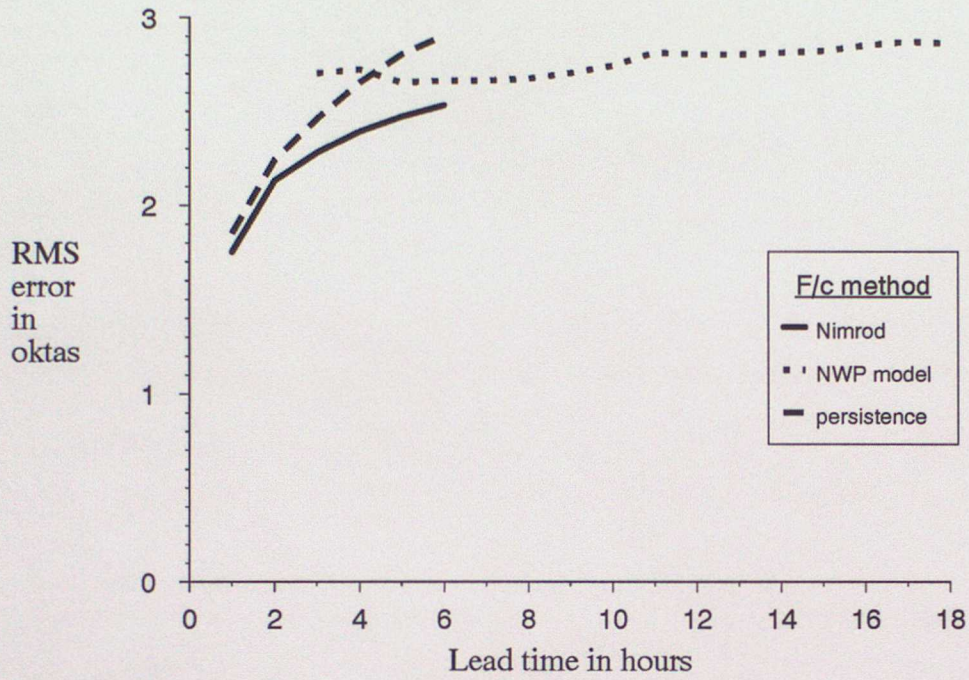

Figure 10 RMS errors in total cloud cover forecasts vs analysis
September 1996 - September 1997

8. Visibility

Variables

The required variables are: mean & worst visibility, and probability of visibility below 200m. The first of these is available and forms the basis of the accuracy measures. A worst visibility product has been developed and is currently computed using a 1% probability. As discussed below, a median (50% probability) product has also been generated and it is proposed that this will replace the "mean". Code to generate the probability of visibility less than 200m has been developed, and allows for other thresholds, currently 1km and 5km. Assessment of these products is in progress.

Forecast Cycle & Interval

Visibility forecasts are output every hour to 6 hours ahead from each forecast, rather than every half hour as requested. Forecasts are updated every hour as requested.

Domain & Resolution

Visibility analyses and forecasts are provided on a 5km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid and mean visibility is reprojected to Polar Stereographic projection. All specifications are satisfied.

Timeliness

The requirement was for visibility analyses and forecasts to be available by 40mins after data time. They were available by T+28 on average and T+32 on 95% of occasions in September 1997.

Accuracy of location of fog (visibility less than 1km)

The mesoscale model shows clear over prediction of fog (figure 11). This is probably due to the method of calculating "mean" visibility. Since the definition works both ways in the analysis - affecting the use of observations as well as the diagnosis of output - there is no corresponding problem with the persistence

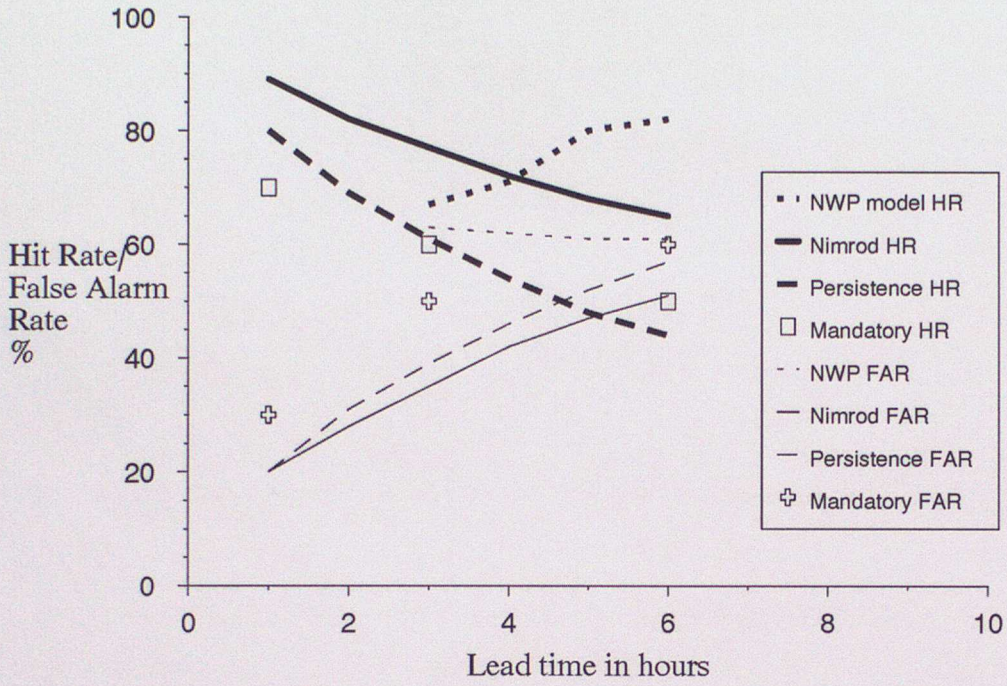

Fig. 11 Accuracy of smoothed areas of visibility < 1km vs analyses
November 1996 - September 1997.

forecast. The merged Nimrod forecast increasingly over predicts at longer ranges as the use of the model increases. The result is that both Nimrod and persistence meet the mandatory criteria. With a substantially higher hit rate and a slightly lower false alarm rate, Nimrod is significantly the better.

Unfortunately this picture is not replicated in the verification against observations (figure 12) which not only shows much reduced accuracy, as one would expect due to representativeness differences, but

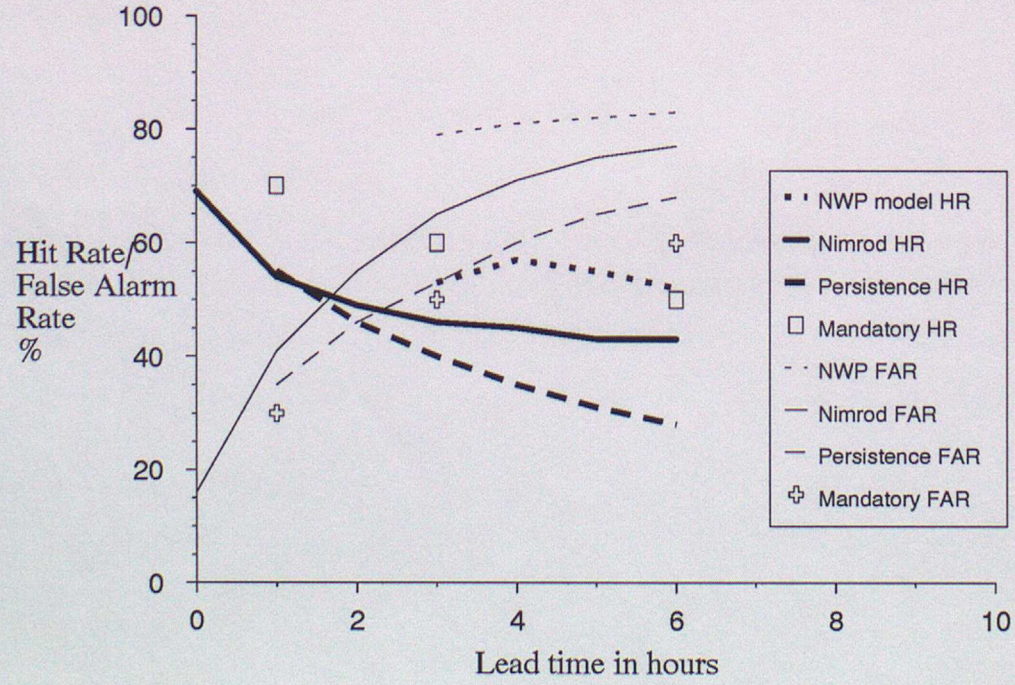

Fig. 12 Accuracy of site specific forecasts of visibility < 1km vs observations
November 1996 - September 1997.

also shows persistence to be a better forecast than Nimrod, due mainly to it having a much lower false alarm rate. Examination of results from the area verification (figure 13), in which pre-defined regions of the British Isles are classified as foggy or not foggy according to the ratio of foggy to non-foggy observations, suggests that the change in absolute accuracy is indeed due to representativeness errors, and that part of the change relative to persistence is also due to this. According to this assessment, the hit rate meets the requirement, the false alarm rate is far too high, and Nimrod and persistence are of roughly equal quality. The differences between the characteristics of figures 11,12 & 13 emphasise the difficulty and resulting uncertainty in verifying visibility products.

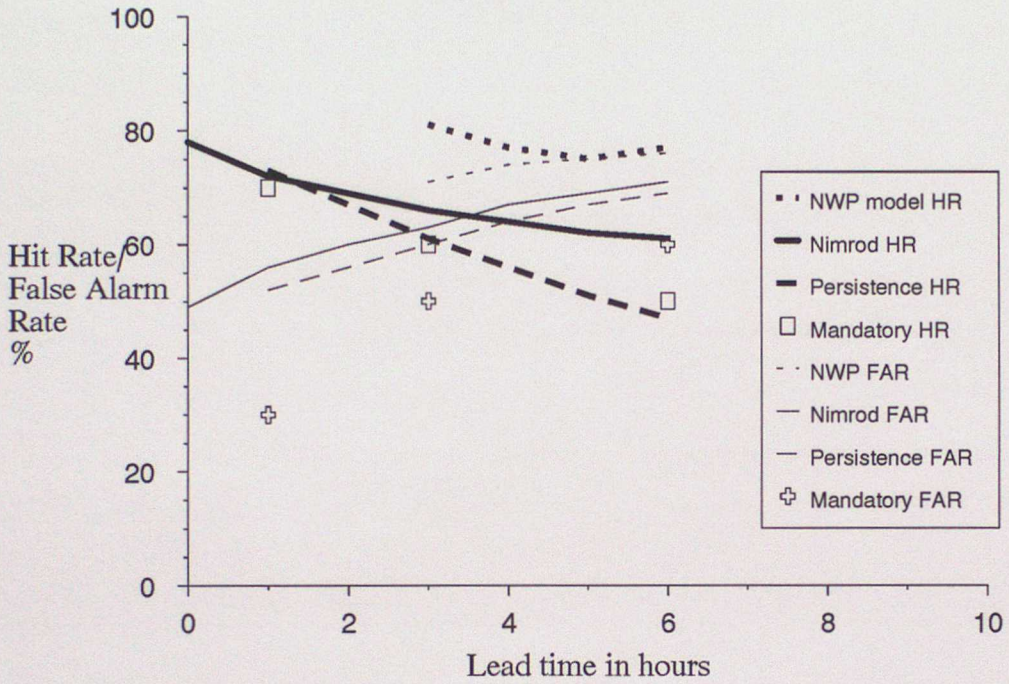

Fig. 13 Accuracy of regional forecasts of visibility < 1km vs observations November 1996 - September 1997.

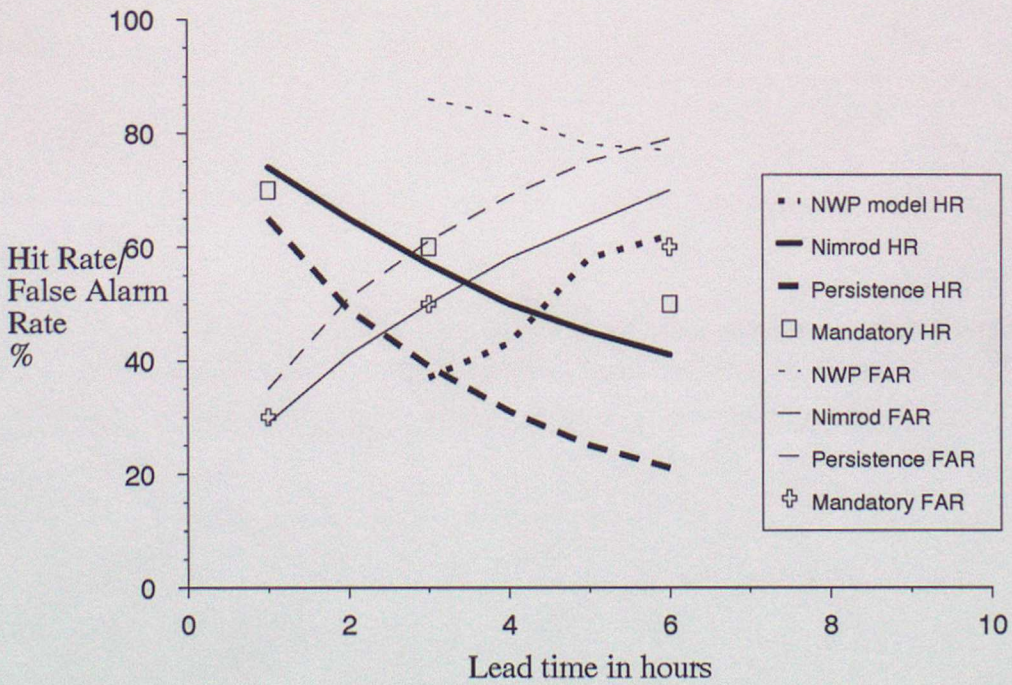

Fig. 14 Accuracy of smoothed areas of visibility < 200m vs analyses November 1996 - September 1997.

Accuracy of location of thick fog (visibility less than 200m)

Areas of visibility less than 200m rarely reach the scale of the smoothing operator so the smoothed verification results may be misleading. The comments made above concerning the quality of the analysis should also be noted here.

The results in figure 14 show that the mesoscale model has no skill at this threshold. Persistence and Nimrod show some skill at very short lead times, Nimrod performs significantly better than persistence, reaching the mandatory acceptance criterion at T+1, and nearly doing so at T+3.

Accuracy of Visibility

No requirement was specified in the project for this measure. However, it is noted that the ICAO requirement for accuracy of site specific visibility forecasts in TAFs is that 75% of forecasts should be within the greater of 200m or 30% of actual.

Figure 15 shows the RMSF error measured against analyses at full resolution where analysed or forecast visibility is less than 5km. Higher values are capped at 10km. As noted already in the precipitation section, verification at full resolution underestimates true accuracy because of aliasing and because pixel scale noise tends to dominate. However, any averaging procedure would be ambiguous (as it was for cloud base) and so none was adopted.

The figure emphasises the poor mesoscale model performance. However, it should be noted that "mean" visibility is not normally a mesoscale model products; only probability of fog, which is believed to be much more reliable. The method of calculation means that a mean 1km visibility corresponds to a probability of , say, 10% that visibility will be below 1km. Subjective experience suggests that the best match of model to observations is obtained for a probability of about 30%.

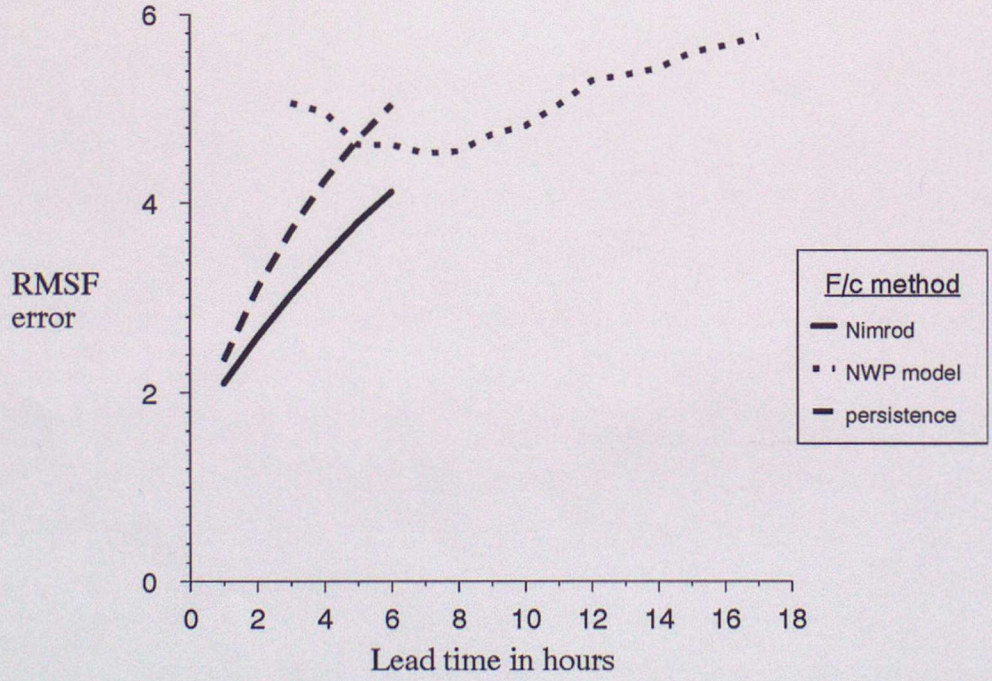

Figure 15 Fractional errors in visibility forecasts vs analysis
November 1996 - September 1997

The persistence error rises very rapidly, emphasising the low temporal correlation of visibility. Nimrod improves on persistence, but not dramatically so - an improved method of interpreting the model visibility is needed in order to make further progress. The intention is to change from use of mean visibility to the median (50% probability)

In terms of the ICAO requirement, Nimrod falls far short, achieving only 20% (63%) at T+1, falling to 10% (20%) at T+6 relative to observations (analyses).

Accuracy of Worst Visibility

Verification of the worst visibility product is limited to hit rate and false alarm rate for thresholds of 200m and 1km, measured against point observations. The aim is, of course, to obtain a 100% hit rate, while keeping the false alarm rate as low as possible. Results for November 1996 to September 1997 show that for the 1km (200m) threshold, the hit rate falls from 88% (87%) at T+1 to 75% (73%) at T+6, while the false alarm rate is approximately constant at about 86% (94%).

Accuracy of low cloud & visibility determined subjectively in NMC

During February and March 1997, the Nimrod cloud base > 4 oktas and visibility products were used to prepare TREND forecasts for Heathrow and Gatwick, in parallel with operational production. The verification concentrates on situations where changes occur or are forecast, and assesses the extent to which observed and forecast changes are consistent. A total of 639 forecasts were verified.

The results are considered in two ways here:

(i) Percentage of occasions on which each method was the better:
Percentage of forecasts for which Nimrod was better than operational: 8.6%
Percentage of forecasts for which Nimrod was worse than operational: 27.1%

(ii) Comparison of the numbers of "good" and "bad" forecasts for each method:
Scores 0,1 include all bad forecasts.
Scores 2,3 include correct forecasts of no change, and slight over forecasting of change.
Scores 4,5 include all correct forecasts of change.

Score	Nimrod	Operational	
0	9%	5%	(change f/c, change obs in opp sense; or no change f/c, >1 limit obs)
1	21%	11%	(change f/c, <1limit obs in opp sense; or no change f/c, 1 limit obs)
2	9%	9%	(One limit change f/c, change obs less than limit)
3	52%	55%	(No change obs & f/c; or >1 limit change f/c, 1 limit obs)
4	6%	14%	(Change of one category f/c & obs)
5	3%	6%	(Change >1 category f/c & obs)

It can be concluded that the operational TREND forecasts are significantly superior to Nimrod forecasts, but that it would be acceptable to use 70% of Nimrod forecasts unaltered. The percentage of unsafe forecasts is 30% for Nimrod and 16% for operational products.

9. Temperature

Temperature is not a required variable from Nimrod. However, since it is produced routinely as part of the visibility forecasting process, and since it provides a significant improvement on the model product in resolution, timeliness and accuracy, it is offered as a product that might have a customer. Dew point and other humidity variables are also available under the same conditions as temperature.

Forecast Cycle & Interval

Temperature forecasts are output every hour to 6 hours ahead and updated every hour.

Domain & Resolution

Temperature analyses and forecasts are provided on a 5km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid.

Timeliness

Temperature forecasts are produced at the same time as visibility forecasts and were available by T+28 on average and T+32 on 95% of occasions in September 1997.

Accuracy

The mesoscale model has a good record of screen temperature prediction, its largest errors being on clear winter nights and summer afternoons. The routine model verification against observations underestimates the model's accuracy, since it does not allow for representativeness errors. Figure 16 shows errors computed against Nimrod analyses.

Nimrod performs very well compared with the model, decreasing errors at all ranges (bearing in mind that the T+6 forecast uses model forecasts of T+9 -T+15). Note the rapid increase in persistence error (persistence of the T+0 value, not of the previous day's value).

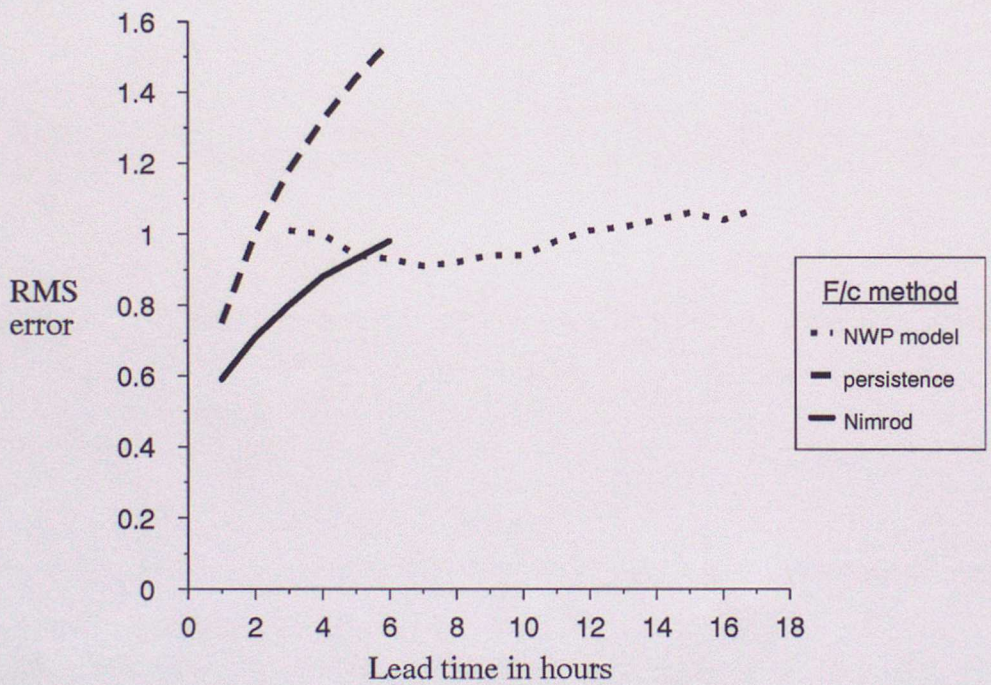

Figure 16 RMS errors in temperature forecasts vs analysis
November 1996 - September 1997

Comparison against point observations shows that the analysis fits observations to a RMS error of 0.65K, probably rather closer than is justified. Forecast verification against observations increases from 0.94K at T+1 to 1.25K at T+6, a significant improvement on the model values of around 1.25K

Analysis of accuracy vs observations as a function of time of day, shows that the improvement over the mesoscale model during the afternoon is short lived, lasting only up to about T+3. However, Nimrod delivers maximum benefit in the early morning, particularly around dawn, when a gain of over 0.1K in 1.2K (~8%) is achieved even at T+6. As an example, Nimrod temperature forecasts issued at the same time as the mesoscale model midnight run (~0230UTC) have had an RMS error over the past year of

0.97°C (model = 1.17°C) at 04UTC, increasing to 1.07°C (1.19°C) at 06UTC, and to 1.14°C (1.19°C) at 08UTC.

The results for dew point are consistently better than the model at all forecast times throughout the day with a small benefit at T+6 in the afternoon, increasing to a maximum of about 0.1°C at 06UTC.

10. Wind speed & direction

It is currently considered that the accuracy of direct model products is sufficient that a more frequently updated Nimrod product is not justified. However, the techniques to provide a more accurate and timely product have been developed for the army sponsored CMETS project, and could be implemented if required.

Forecast Cycle & Interval

Wind forecasts are output every hour to 18 hours ahead and are updated every 6 hours rather than the hourly cycle requested.

Domain & Resolution

Wind products are available from Nimrod on a 15km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are available on the UK National Grid.

Levels

Wind products are currently available from Nimrod at 10m only. The code to interpolate 50m wind products has been developed and awaits a customer before implementation.

Timeliness

Wind products from the mesoscale Unified Model are available from Nimrod about 3.5 hours after data time.

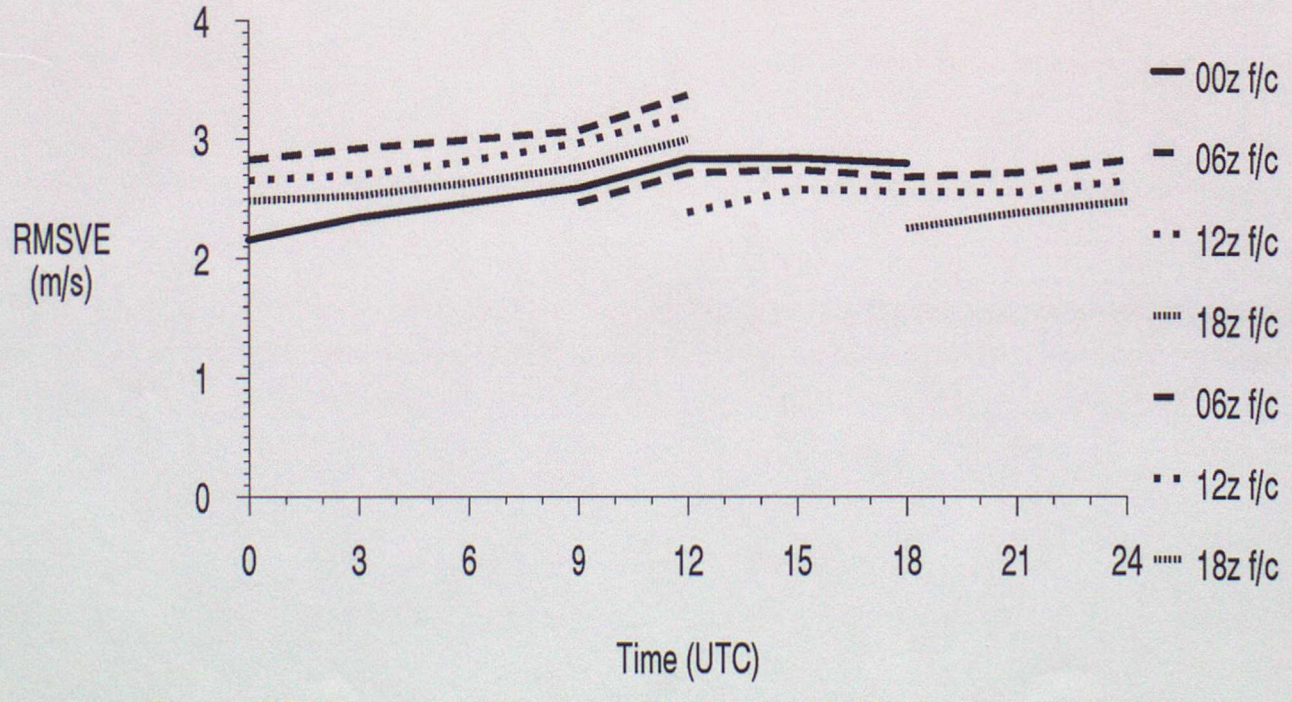

Figure 17 RMS vector errors in mesoscale model wind forecasts vs observations
October 1996 - September 1997

Accuracy of 10m wind

Mesoscale model surface wind is verified against observations in the routine operational verification scheme administered by (FS4). Since the acceptance criteria were for comparison with analyses, it is necessary to adjust them for representativeness error (3kn for wind speed; 4.2kn for wind vector), giving mandatory (target) criteria of 5kn (4.2kn) for speed and 7.3kn (5.9kn) for vector when compared with observations. Figure 17 shows the RMS vector error as a function of time-of-day for each of the forecast runs. All achieve the adjusted target criterion ($2.9ms^{-1}$) up to T+12. The accuracy of wind speed also meets the adjusted target criterion.

A further mandatory (target) criterion was specified in terms of strong winds of force 7 or more, being matched within 2(1) hour on 75% (80%) of occasions. This has not been measured for mean winds, but see the section on gust forecasts below.

Accuracy of 50m wind

There is currently no identified customer for this product so it is not currently generated. Evidence from LAM verification vs radiosondes, and from a defence study on accuracy of artillery forecasts, suggests that the results should approach target quality (RMSVE < 6kn).

11. Wind gust

This product is generated from consideration of two sources of wind variability: boundary layer turbulence and convective downdraughts. The first is solely model information, but the second depends on the precipitation distribution and so can be updated after each precipitation forecast. The product is a maximum gust, but does not take account of speed-up by features in the topography.

Whereas gust observations relate to the maximum wind speed observed at a point over a period, normally one hour, the Nimrod product is the maximum over an area 15km square, of the one hour maximum at each point. The resulting increase in the number of effective samples of the wind speed distribution means that the Nimrod maximum gust estimate should exceed that observed, especially in strongly varying forcing such as convective downdraughts. On the other hand, it should provide a more reliable upper bound to the gusts observed.

Forecast Cycle & Interval

Wind gust forecasts are output every hour to 6 hours ahead and are updated every hour.

Domain & Resolution

Wind gust diagnoses and forecasts are provided on a 15km grid covering the British Isles, surrounding waters, including the North Sea, and nearby parts of Europe. Data are currently available on the UK National Grid. The projection and format of the product will be determined when a customer wishes to receive it.

Timeliness

The wind gust product is generated at the same time as precipitation type and was available by T+17 on average and T+21 on 95% of occasions during September 1997.

Accuracy

The required mandatory (target) criterion was specified in terms of gusts of force 7 or more, being matched within 2(1) hour on 75% (80%) of occasions. This is taken to mean that given an observation of force 7 or more, a forecast of force 7 or more should be provided for a time within 2 hours, i.e.: that

the hit rate for force 7 or more is at least 75%. The results in figure 4 show a contingency table of gusts reported vs gusts diagnosed without allowance for any time delay. Comparisons are only made when a gust is reported.

There is a clear correlation between observations and forecasts and an overall hit rate (false alarm rate) of 77%(49%) for a threshold of force 7, and 65%(69%) for a threshold of force 8. In winter non-convective conditions, the forecasts are too high on average, whereas in summer convective conditions, the results are unbiased and generally rather poor. The winter bias is deliberate, ensuring that gusts at coastal stations, for instance, are not underestimated. The summer performance is affected by a recently discovered error. Further study of performance will be required following correction of this error .
Although these results refer to T+0, deterioration at longer times is minimal (less than 2% at T+6).

	f/c F5	f/c F6	f/c F7	f/c F8	f/c >F8
obs F5	36.9%	6.9%	3.0%	0.8%	0.2%
obs F6	7.0%	10.6%	8.4%	3.5%	0.7%
obs F7	0.9%	3.3%	5.2%	4.2%	1.4%
obs F8	0.1%	0.5%	1.4%	2.0%	1.4%
obs >F8	0.1%	0.1%	0.2%	0.5%	0.7%

Table 4. Contingency table of gust observations and diagnoses
January - September 1997

12. System Maintenance

Code

The code has all been reviewed to ensure that it conforms to the Nimrod code and internal documentation standards.

Code is maintained using the RCS code management system which ensures that past versions can always be retrieved.

Change Control

A change control system is in force based on a dual purpose form which is used to record both faults and changes.

Changes to algorithms are approved via circulation of a trial report to user representatives who indicate their acceptance or otherwise on a standard PROMMs form.

Documentation

Management papers are available on the project files which will be retained by the project manager for a period to be determined after project completion.

Working papers and reports describing the algorithms and their performance are retained in box files by the librarian. They will be retained by the project manager for a period after project completion.

Final versions of the algorithms, results from trials, and system design & performance have been published as internal technical reports &/or in external journals. Reference to these papers is made in

End Stage reports and in the Project Overview document, which are available from the project manager.

On-line reference

The Nimrod real-time monitor provides extensive facilities for management of the system and for reference. The monitor is accessed via the intranet using a web browser. The top-level "warnings" page is used by HQMC and the system manager to provide an instant alert if data are missing or if the system is behaving abnormally. At the next level, most of the current products are available for browsing, together with information on times of recent data arrival and product dissemination. Log output from the most recent run of many algorithms is also accessible. A third level of data is concerned with statistics of recent performance including monthly and annual accuracy, daily and monthly timings, and a list of system changes. Finally, there are reference pages available giving the structure of the system, file and archive lists, a hardware description, structured code listings and sample case studies. Together, these facilities add up to probably the most comprehensive on-line support for any operational system in the Met. Office.

13. Conclusions & Recommendations

General

(i) The current availability, forecast and update cycles, and timeliness are accepted.

(ii) There is a requirement for a msl pressure nowcast for use in automated QNH forecasting. A project to develop such a product should be put forward next year as part of the review of NMC automation requirements.

Precipitation rate & accumulation

(i) Precipitation rate and accumulation have already been accepted as operational.

(ii) Future development will be concentrated on improved radar interpretation, supported by EA funding if possible.

(iii) Improved convective precipitation structure in model output will be developed through the UM convection diagnosis project.

(iv) Accumulations will be provided to Horace, Nimbus & other users at appropriate priority.

(v) A probability of precipitation product will be implemented to assist users in interpreting the radar actual. It will initially be used to provide a second radar actual, retaining only high probability precipitation echoes.

(vi) A regional 2km resolution composite will be developed in 1998.

Precipitation Type

(i) Precipitation type is already operational to external customers.

(ii) An improved diagnosis of freezing rain is being developed (target availability 12/1997).

(iii) A rime product for commercial, aviation and defence users will be implemented in early 1998.

(iv) A more detailed verification of hail diagnosis quality will be completed in 1998.

Snow probability

(i) Probability of snow is accepted as operational and will be delivered to users at appropriate priority.

(ii) Probabilities of other precipitation types will be developed as a means of improving precipitation type. Users will consider whether they meet a need individually as they become available.

Lightning Rate

(i) Results now show a high probability of lightning detection, and a strong relationship between forecast flash rate and observed lightning risk. The product is accepted as operational.

(ii) Lightning risk will be delivered to Horace, Nimbus & other users at appropriate priority.

(iii) Performance will be reviewed following the ATD upgrade.

Cloud

(i) Objective criteria are met except for cloud below 1000'. Subjective assessment of TREND forecasts indicates that the product can be a useful first guess for subjective forecasting. It is accepted as operational.

(ii) The model bias correction has been shown to improve the performance below 1000'. It will be implemented at the earliest opportunity.

(iii) The need for further work on low cloud will be considered after analysis of this winter's results.

Visibility

(i) Objective criteria are met at all lead times for fog but only at T+1 for thick fog. These results are subject to reservations about the differences between scores against analyses and scores against observations. Subjective assessment of TREND forecasts indicates that the product can be a useful first guess for subjective forecasting. It is accepted as operational.

(ii) The revised visibility analysis and forecast algorithms will be implemented at the earliest opportunity, including provision of probability of thick fog product.

(iii) Further work on visibility will be focussed on the site-specific forecast model.

Temperature

Temperature and dew-point results are of high quality but do not currently meet a need, so should not be disseminated.

Wind speed & direction

The quality of mesoscale model output meets the target users requirements for mean wind. However, a need exists for a wind nowcasting scheme to provide short range warning of occasions when strong winds are under forecast by the model. Development of such a scheme should be undertaken in conjunction with the msl pressure nowcasting scheme.

Wind gust

The current quality of wind gust forecasts does not meet a need. Further work to improve the quality may be undertaken in conjunction with development of the wind nowcasting scheme.